The Annals of Statistics
1996, Vol. 24, No. 1, 41–63

ON THE ASYMPTOTIC PROPERTIES OF A FLEXIBLE HAZARD ESTIMATOR[1]

BY ROBERT L. STRAWDERMAN AND ANASTASIOS A. TSIATIS

University of Michigan and Harvard University

Suppose one has a stochastic time-dependent covariate $Z(t)$, and is interested in estimating the hazard relationship $\lambda(t|\bar{Z}(t)) = \omega(Z(t))$, where $\bar{Z}(t)$ denotes the history of $Z(t)$ up to and including time t. In this paper, we consider a model of the form $\exp(s_n(Z(t)))$, where $s_n(Z(t))$ is a spline of finite but arbitrary order, and investigate the behavior of the maximum likelihood estimator of the hazard as the number of knots in the spline function increases with the sample size at some rate $k_n = o(n)$. For twice continuously differentiable $\omega(\cdot)$, we demonstrate that the difference between the estimator $\exp(s_n(\cdot))$ and $\omega(\cdot)$ goes to 0 in probability in sup-norm for any $k_n = n^{\phi}$, $\phi \in (0,1)$. In addition, if $\phi > 1/5$, then $\exp(\hat{s}_n(Z(t))) - \omega(Z(t))$, properly normalized, is asymptotically standard normal. A large-sample approximation to the variance is derived in the case where $s_n(\cdot)$ is a linear spline, and exposes some rather interesting behavior.

1. Introduction. Suppose one is interested in estimating the conditional hazard function $\lambda(t|\bar{Z}(t))$, where $\bar{Z}(t)$ is the history of a stochastic time-dependent covariate $Z(u)$ for $u \in [0,t]$. An interesting and important class of problems in health-related research is when the hazard function $\lambda(t|\bar{Z}(t)) = \omega(Z(t))$; that is, the hazard, taken conditionally on the observed covariate history $\bar{Z}(t)$, is a function of the current value of the covariate alone. Such a relationship might be expected to occur if observed differences in an outcome variable (such as survival) are primarily mediated through differences in the time-dependent covariate $Z(t)$. For example, in AIDS research, an individual's prognosis is measured by the level of destruction within the immune system. This may be measured to some extent by the observed level of that individual's $CD4^+$-lymphocyte count. Thus, one might expect the above relationship to hold if an infected individual's $CD4^+$ count level is the primary mechanism by which their ultimate survival is determined. We shall consider such models to be the "truth" throughout this article, meaning that it is assumed throughout that the underlying hazard function being estimated has the functional form $\omega(Z(t))$.

Since this functional relationship is generally unknown, a flexible family of models may be used. One possibility is $\omega_n(Z(t), \beta) = \exp(\beta'\mathbf{f}(Z(t)))$, where $\mathbf{f}(\cdot)$ defines a B-spline basis for a given set of knots. The parameters of this

Received July 1992; revised June 1995.

[1]Supported in part by NCI Grant CA-51962 and NIAID grant AI-31789-01.

AMS 1991 *subject classifications*. Primary 62F12, 62G05; secondary 60G44, 62M99, 62P10.

Key words and phrases. Censoring, hazard function, martingale, spline, survival analysis, time-dependent covariate.

model are easily estimated via maximum likelihood, and the standard asymptotic properties of the resultant maximum likelihood estimate are easily derived under reasonably mild regularity assumptions, the most stringent being that the model is a correct representation of the truth [Strawderman (1992)].

Usually, this last assumption will be violated, at least to some degree. Splines are often used in statistical modeling because of their ability to approximate unknown functions. In particular, as the sample size grows, increasing the number of knots in a spline function at some lesser rate generally allows the spline-based estimator to get uniformly closer to the true function at some rate that depends primarily on the maximum difference between adjacent knots. There have been a significant number of papers concerned with proving such results for various statistical models; see, for example, Speckman (1985) and Portnoy (1988). Stone (1980, 1982, 1986, 1990, 1991, 1994) has made numerous contributions to this literature, most recently with the development of a unified framework in which consistency and optimal L_2 rates of convergence for spline estimators in regression and density estimation problems can be established. Kooperberg, Stone and Truong (1995b) extend the results of Stone (1994) to hazard regression under noninformative censoring, but do not consider time-dependent covariates.

In this paper, the behavior of $\omega_n(\cdot, \hat{\beta})$ is investigated for splines of arbitrary but fixed order. The general problem is described in detail in Section 2, where we give all relevant definitions and assumptions. In Section 3, we prove that $\omega_n(\cdot, \hat{\beta}) - \omega(\cdot)$ converges in probability to 0 in sup-norm as the number of knots grows at a lesser rate than the sample size. This is done by first proving the existence of a deterministic sequence of "least-false parameters" [Hjort (1992)], say $\{\beta^*\}$, such that $\|\omega_n(\cdot, \beta^*) - \omega\|_\infty \to 0$. Then, we show that $\|\omega_n(\cdot, \hat{\beta}) - \omega\|_\infty \to_P 0$ by demonstrating the existence of the sequence of MLEs $\{\hat{\beta}\}$ such that $\|\hat{\beta} - \beta^*\|_\infty \to_P 0$. Existence and consistency are established simultaneously in each case using a modification of the inverse function theorem [cf. Foutz (1977)]. In Section 4, it is shown via the martingale central limit theorem that $\log \omega_n(\cdot, \hat{\beta}) - \log \omega(\cdot)$ is asymptotically standard normal under proper normalization. Some remarks on useful extensions of these results and rates of convergence are given following the proof. Appendix A contains lemmas necessary for completing the proofs of these results. In Appendix B, we derive a large-sample approximation to the asymptotic variance of $\log \omega_n(\cdot, \hat{\beta})$ for the case of linear B-splines. To our knowledge, explicit derivations of the form of the asymptotic variance for spline-based estimators and such convergence results in the presence of stochastic time-dependent covariates have not previously appeared in the literature.

2. Description of the problem. In a typical survival analysis problem, the data consist of triplets $(X_i, \Delta_i, \overline{Z}_i(X_i))$, $i = 1, \dots, n$, where each is an i.i.d. copy of a random triplet $(X, \Delta, \overline{Z}(X))$ defined on some probability space $\{\Omega, \mathscr{F}, Q\}$. The random variable $X = \min(T, C)$, where T and C, respectively, denote random failure and censoring times, and $\Delta = I\{T \le C\}$ is the failure time indicator variable. The covariate $Z(\cdot)$ is assumed to be a continuous-time stochastic process with state space Θ, where Θ is a closed finite interval on $\mathbb{R}$, and $\bar{Z}(X)$ represents the history of this process to X, or $\{Z(u),\ u \in [0, X]\}$. For simplicity, Θ is assumed to be time-independent. However, with some modification and additional assumptions, the results extend in a straightforward manner to the case where the support of $Z(\cdot)$ varies with t.

Throughout, $Z_i(\cdot)$ shall be treated as an ancillary time-dependent covariate [Kalbfleisch and Prentice (1980), Section 5.3]. This assumption requires the entire path of an individual's covariate to be predetermined in the sense that the covariate path may influence, but may not be influenced by, the failure process. For example, suppose the covariate path for each individual follows the growth curve model $Z_i(t) = \gamma_{0i} + \gamma_{1i}t$, where $\gamma_i = (\gamma_{0i}, \gamma_{1i})'$ comes from some underlying bivariate distribution (e.g., bivariate normal) whose parameters do not depend on the failure time distribution for T. Then, given γ_i, the entire history of the covariate path for that individual is predetermined even though we may not be allowed to observe it in its entirety.

Without loss of generality, we assume that observation takes place for $t \in [0,1]$. For technical reasons, we additionally require that (i) for each $t \in [0,1]$, the joint density of T and $Z(t)$ is at least twice continuously differentiable and bounded away from 0 and ∞ on $[0,1] \times \Theta$, and (ii) $\Pr\{T > 1|\bar{Z}(1)\} > 0$. One implication of assumption (i) is that the marginal density of $Z(t)$, denoted $h_Q(\theta; t)$, satisfies $M_1^{-1} \le h_Q(\theta; t) \le M_1$ on $[0,1] \times \Theta$ for some positive finite constant M_1. We assume that censoring is noninformative, and that for all covariate paths the distribution function of the censoring time C is continuous on $[0,1)$ and the event $\{C \le 1\}$ has probability equal to 1. Note that these assumptions imply that the distribution of the observed survival time X satisfies $\Pr\{X \in [0,1]\} = 1$, where X may take on the value 1 with positive probability regardless of $\bar{Z}(1)$. Such an assumption is reasonable, for example, if the observations come from a clinical trial where there is some maximum follow-up time in effect.

In order to effectively use martingale results, we cast the problem in the multiplicative intensity model framework [Aalen (1978)]. For a sample of size n, define for each individual i the 0–1 counting process $N_i(t) = I\ (X_i \le t, \Delta_i = 1)$ and its associated stochastic intensity function $A_i(t) = \omega(Z_i(t))Y_i(t)$, where $Y_i(t) = I\ (X_i \ge t)$ is the usual left-continuous "at-risk" process and $\omega(\cdot)$ is a deterministic twice-continuously differentiable function. Let the right-continuous filtration $\{\mathscr{F}_t\colon 0 \le t \le 1\}$ be the smallest σ-algebra containing all of the information on failure times, censoring and covariate histories up to time t for all individuals. Specifically, if we let $\mathscr{F}_i^{(i)} = \sigma\{N_i(u), Y_i(u), \bar{Z}_i\colon 0 \le u \le t\}$, then $\mathscr{F}_t = \vee_i \mathscr{F}_i^{(i)}$. We require the paths of $Y_i(u)$, $Z_i(u)$ and $A_i(u)$ to be predictable given $\mathscr{F}_t$.

As an estimator for $\omega(Z(t))$, we consider the model

$$(1) \qquad \omega_n(Z(t), \beta) = \exp(\beta'\mathbf{f}(Z(t))),$$

where $\mathbf{f}(\cdot) = (f_1(\cdot), \dots, f_{k_n}(\cdot))'$ is the usual normalized B-spline basis for the space of mth-order polynomial splines with knots $\Xi_n = \{\tau_i\}_1^{k_n+m}$ defined on Θ

[cf. de Boor (1978)]. For each n, we assume that Ξ_n forms an extended uniform partition of Θ with mesh $\overline{\Delta}_n = \max_i\{\tau_i - \tau_{i-1}\} = O(k_n^{-1})$ and τ_m and τ_{k_n} fixed at the boundaries of Θ [cf. Schumaker (1981)]. The basis functions $f_r(\theta)$, $r = 1, \dots, k_n$, are nonnegative on Θ, positive on an interval of length at most $m\overline{\Delta}_n$ and sum to unity for all $\theta \in \Theta$; let the resulting space of splines be denoted by $\mathscr{S}_m(\overline{\Delta}_n)$. For notational simplicity, the dimension of all matrices and vectors are assumed to depend on k_n unless otherwise specified.

The following definitions of sup-norm are used throughout:

1. For any function $r(\mathbf{w})$, $\mathbf{w} \in \mathscr{W}$ such that r: $\mathscr{W} \to \mathbb{R}$, define $\|r\|_\infty = \sup_{w \in \mathscr{W}}|r(\mathbf{w})|$.
2. For any $k_n \times 1$ vector $\mathbf{x}$, define $\|\mathbf{x}\|_\infty = \max_{g=1,\dots,k_n}|x_g|$.
3. For any $k_n \times k_n$ matrix $\mathbf{Y}$, define $\|\mathbf{Y}\|_\infty = \max_{g=1,\dots,k_n}\sum_{g'=1}^{k_\eta}|Y_{gg'}|$.

We note that for a given B-spline basis $\mathbf{f}(\cdot)$ and any conformable vector $\mathbf{x}$, it follows directly from the properties of the B-spline basis that $\|\mathbf{x}'\mathbf{f}\|_\infty \le \|\mathbf{x}\|_\infty$. This property is used in many of the upcoming proofs.

The relevant partial log-likelihood for β based on (1) is

$$l(\beta) = \sum_{i=1}^{n} \int_0^1 \left[\log(\omega_n(Z_i(u), \beta))\, dN_i(u) - \omega_n(Z_i(u), \beta) Y_i(u)\, du\right],$$

with associated (normalized) score vector

$$S_n(\beta) = \frac{k_n}{n} \sum_{i=1}^{n} \int_0^1 \mathbf{f}(Z_i(u))\left[dN_i(u) - \omega_n(Z_i(u), \beta) Y_i(u)\, du\right].$$

If we define the compensated counting process

$$M_i(t) = N_i(t) - \int_0^t \omega(Z_i(s)) Y_i(s)\, ds, \tag{2}$$

then $S_n(\beta)$ may be rewritten as

$$S_n(\beta) = \frac{k_n}{n} \sum_{i=1}^{n} \int_0^1 \mathbf{f}(Z_i(u)) \times\left[dM_i(u) + \left[\omega(Z_i(u)) - \omega_n(Z_i(u), \beta)\right] Y_i(u)\, du\right], \tag{3}$$

which is a stochastic integral with respect to a martingale process plus a remainder term; we note that $S_n(\beta)$ is defined conditionally upon the $\overline{Z}_i$, $i = 1, \dots, n$. The maximum likelihood estimate $\hat{\beta}$ is found by solving the equation $S_n(\beta) = 0$ for β. Similarly, the least-false parameter β^* for a sample of size n is defined as the solution to $E[S_n(\beta)] = 0$, where $E[\cdot]$ denotes the expectation with respect to the true probability distribution of the data [cf. Hjort (1992)]. Letting $S_n^*(\beta) = E[S_n(\beta)]$, we see by (2) and (3) that

$$S_n^*(\beta) = k_n \int_0^1 \int_\Theta \mathbf{f}(\theta)\left[\omega(\theta) - \omega_n(\theta, \beta)\right] p(u, \theta) h_Q(\theta; u)\, d\theta\, du, \tag{4}$$

where $p(u, \theta) = \Pr\{X \geq u | Z(u) = \theta\}$. Under the assumptions given earlier, $p(u, \theta)$ is well defined and positive on $[0, 1] \times \Theta$.

Let

$$H_n(\beta) = -\frac{\partial}{\partial\beta} S_n(\beta) \tag{5}$$

and

$$\mathscr{I}_n(\beta) = k_n \int_0^1 \int_\Theta \mathbf{f}(\theta)\mathbf{f}'(\theta)\,\omega_n(\theta, \beta)\,p(u, \theta)\,h_Q(\theta; u)\,d\theta\,du \tag{6}$$

be the negative of the first derivatives of (3) and (4), respectively. Note that (6) is also the expectation of (5) under the true probability distribution of the data.

With these definitions and assumptions, we investigate the asymptotic behavior of $\hat{\beta}$, β^* and various quantities based on them. Specifically, in Section 3, it is shown that both $\hat{\beta}$ and β^* exist and that $\|\omega_n(\cdot, \beta^*) - \omega\|_\infty \to 0$ and $\|\omega_n(\cdot, \hat{\beta}) - \omega\|_\infty \to_P 0$ as $n \to \infty$. For this proof, we only require that $k_n = n^\phi$ for $\phi \in (0, 1)$. In Section 4, we prove that $\log \omega_n(\nu, \hat{\beta}) - \log \omega(\nu)$ for $\nu \in \Theta^o$ (the interior of Θ), properly normalized, is asymptotically standard normal when $k_n = n^\phi$ for $\phi \in (\frac{1}{5}, 1)$. This proof involves an application of the martingale central limit theorem, and arises out of the decomposition of the score vector as given by (3).

3. Consistency.

The main result is given in Theorem 1, and its proof constitutes Sections 3.1 and 3.2.

THEOREM 1. *Let $\hat{\beta}$ and β^* denote the solutions to $S_n(\beta) = 0$ and $S_n^*(\beta) = 0$, respectively. Then, as $n \to \infty$ and for $k_n = n^\phi$ where $\phi \in (0, 1)$:*

(a) *β^* exists and $\|\omega_n(\cdot, \beta^*) - \omega\|_\infty \to 0$;*
(b) *$\hat{\beta}$ exists with probability going to 1 and $\|\omega_n(\cdot, \hat{\beta}) - \omega\|_\infty \to_P 0$.*

The inverse function theorem [IFT; see Rudin (1964)] will be used to prove both (a) and (b). The proof of (a), done in Section 3.1, requires a version of the IFT modified for use with sup-norm. The necessary conditions are incorporated as part of the proof. The proof of (b) is given in Section 3.2 and employs a stochastic version of the same. The idea for using the IFT is taken from Foutz (1977), who demonstrated how it could be used to simultaneously prove existence and consistency in finite parameter problems. Using sup-norm, Strawderman and Tsiatis (1995) formulate a modification of Foutz's result which is appropriate for parameter spaces of expanding dimensions, and the result is given as Lemma 1 in Section 3.2.

3.1. *Proof of Theorem* 1(a). We shall later show that for n sufficiently large there exist solutions $\{\beta^*: S_n^*(\beta^*) = 0\}$. Assuming this to be the case, we desire to prove that

$$\|\omega_n(\cdot, \beta^*) - \omega\|_\infty \to 0. \tag{7}$$

By the triangle inequality, it is easily seen that for any other sequence $\{\beta^{**}\}$,

$$\| \omega_n(\cdot, \beta^*) - \omega\|_\infty \le \| \omega_n(\cdot, \beta^{**}) - \omega\|_\infty + \| \omega_n(\cdot, \beta^*) - \omega_n(\cdot, \beta^{**}) \|_\infty,$$

and the problem may be reduced to proving that both quantities on the right-hand side converge to 0 as $n \to \infty$. Since $\omega(\cdot)$ is bounded and continuous with bounded and continuous first and second derivatives, standard approximation theory results yield for sufficiently large n, say $n > N_1 \gg m$, the existence of coefficients β^{**} which may be defined through a set of bounded linear functionals of the form $\lambda_i \log \omega$, $i = 1, \dots, k_n$, such that $\|\log \omega_n(\cdot, \beta^{**}) - \log \omega\|_\infty = O(\overline{\Delta}_n^2)$ and $\|\log \omega_n(\cdot, \beta^{**})\|_\infty < M_2$, where $M_2 < \infty$ depends only on $\|\log \omega\|_\infty$ and the order of the spline m [cf. de Boor (1978), Chapter 12]. It follows by continuity that

$$\| \omega_n(\cdot, \beta^{**}) - \omega\|_\infty = O\big(\overline{\Delta}_n^2\big) \tag{8}$$

and

$$\| \omega_n(\cdot, \beta^{**}) \|_\infty \le M_2, \tag{9}$$

with M_2 appropriately redefined. The proof of (a) now follows if $\|\omega_n(\cdot, \beta^*) - \omega_n(\cdot, \beta^{**})\|_\infty \to 0$ and the existence of β^* can be demonstrated.

These results may be obtained simultaneously using the IFT. Suppose that $S_n^*(\beta)$ satisfies the conditions of the IFT in an ε-neighborhood about β^{**}, say U_ε, and the value 0 lies interior to a set of proportional size centered at $S_n^*(\beta^{**})$ which is itself contained in the image set $S_n^*(U_\varepsilon)$. Then, the existence of a locally unique solution $\{\beta^*: S_n^*(\beta^*) = 0\}$ such that $\|\beta^* - \beta^{**}\|_\infty = O(\|S_n^*(\beta^{**})\|_\infty)$ is guaranteed. More specifically, if we can demonstrate the existence of constants $\varepsilon > 0$, $N_\varepsilon < \infty$ and $M_3 < \infty$ such that, for $n > N_\varepsilon$:

(i) $\|\mathscr{I}_n^{-1}(\beta^{**})\|_\infty \le M_3$,
(ii) $\sup_{\|\beta - \beta^{**}\|_\infty < \varepsilon} \|\mathscr{I}_n(\beta) - \mathscr{I}_n(\beta^{**})\|_\infty \le 1/(2M_3)$,
(iii) $\|S_n^*(\beta^{**})\|_\infty = O(\overline{\Delta}_n^2)$,

then β^* exists for such n and $\|\beta^* - \beta^{**}\|_\infty = O(\overline{\Delta}_n^2)$. The proof of (a) then follows by the continuity of $\omega_n(\cdot, \beta)$ as a function of β.

Using Lemma A.1, it is relatively easy to prove (i). From (6),

$$\mathscr{I}_n(\beta^{**}) = k_n \int_0^1 \int_\Theta \mathbf{f}(\theta)\mathbf{f}'(\theta)\,\omega_n(\theta, \beta^{**})\,p(u, \theta)\,h_Q(\theta; u)\,d\theta\,du. \tag{10}$$

Let $y_n(u, \theta) = \omega_n(\theta, \beta^{**})p(u, \theta)h_Q(\theta; u)$. Then, under the assumptions of Section 2 and the fact that β^{**} is bounded, the function $y_n(u, \theta)$ is continuous and bounded away from 0 and ∞ on $[0, 1] \times \Theta$ for $n > N_2 \ge N_1$. Thus, by Lemma A.1,

$$\left\|\mathscr{I}_n^{-1}(\beta^{**})\right\|_\infty \le M_3 \tag{11}$$

for $n > N_1$ and an appropriately chosen constant $M_3 < \infty$.

In order to prove (ii), we first use the properties of sup-norm to get that

$$\|\mathscr{I}_n(\beta) - \mathscr{I}_n(\beta^{**})\|_\infty \le \left\| k_n \int_0^1 \int_\Theta \mathbf{f}(\theta)\mathbf{f}'(\theta) p(u,\theta) h_Q(\theta;u)\, d\theta\, du \right\|_\infty \times \|\omega_n(\cdot,\beta) - \omega_n(\cdot,\beta^{**})\|_\infty.$$

Lemma A.1 may be used to bound the first term on the right-hand side by a constant $M_4 < \infty$. From (9) we know that $\|\omega_n(\cdot,\beta^{**})\|_\infty \le M_2$. Since $\omega_n(\cdot,\beta)$ is continuously differentiable in β, applying the mean value theorem yields

$$\sup_{\|\beta - \beta^{**}\|_\infty < \varepsilon} \|\omega_n(\cdot,\beta) - \omega_n(\cdot,\beta^{**})\|_\infty \le M_5\varepsilon$$

for a constant $M_5 < \infty$. Hence, setting $\varepsilon \le 1/(2M_3M_4M_5)$, it follows that

$$\sup_{\|\beta - \beta^{**}\|_\infty < \varepsilon} \|\mathscr{I}_n(\beta) - \mathscr{I}_n(\beta^{**})\|_\infty \le \frac{1}{2M_3}.$$

Thus, an $\varepsilon > 0$ exists independently of $n > N_2$ such that (ii) is satisfied. Consequently, for sufficiently large n, $S_n^*(\beta)$ satisfies the requisite conditions needed to apply the IFT in an ε-neighborhood about β^{**}.

To prove (iii), we note from (4) and the properties of sup-norm that

$$\begin{aligned}\|S_n^*(\beta^{**})\|_\infty &\le \left\| k_n \int_0^1 \int_\Theta \mathbf{f}(\theta) p(u,\theta) h_Q(\theta;u)\, d\theta\, du \right\|_\infty \|\omega_n(\cdot,\beta^{**}) - \omega\|_\infty \\ &\le M_1 \left\| k_n \int_\Theta \mathbf{f}(\theta)\, d\theta \right\|_\infty O(\overline{\Delta}_n^2),\end{aligned}$$

where the last part follows from (8) and the fact that $\|h_Q(\cdot;\cdot)\|_\infty \le M_1$ and $\|p(\cdot,\cdot)\|_\infty \le 1$. Since $\|k_n \int_\Theta \mathbf{f}(\theta)\, d\theta\|_\infty = O(1)$, it follows that $\|S_n^*(\beta^{**})\|_\infty = O(\overline{\Delta}_n^2)$. Hence, we may find an N_ε such that $\|S_n^*(\beta^{**})\|_\infty$ is sufficiently small enough to guarantee that 0 lies in the image set $S_n^*(U_\varepsilon)$.

Thus, for $n > N_\varepsilon$, a locally unique solution $\{\beta^*: S_n^*(\beta^*) = 0\}$ exists such that $\|\beta^* - \beta^{**}\|_\infty = O(\overline{\Delta}_n^2)$. Since $\omega_n(\cdot,\beta)$ is continuous, it follows that, for such n,

$$\|\omega_n(\cdot,\beta^*) - \omega_n(\cdot,\beta^{**})\|_\infty = O(\overline{\Delta}_n^2), \tag{12}$$

completing the proof. In conjunction with (9), this also implies

$$\|\omega_n(\cdot,\beta^*)\|_\infty < M_6 \tag{13}$$

for some constant $M_6 < \infty$. □

3.2. *Proof of Theorem* 1(b). A stochastic formulation of conditions (i) to (iii) in Section 3.1 will be needed, and is given below as Lemma 1. The lemma is proved in Strawderman and Tsiatis (1995) and parallels the proof found in Rudin (1964). The matrices $H_n(\beta)$ and $\mathscr{F}_n(\beta)$ are respectively defined in (5) and (6).

Lemma 1. *Let $N^* < \infty$. For $n > N^*$, suppose that $S_n(\beta)$ is a continuously differentiable mapping from $\mathbb{R}^{k_n}$ to $\mathbb{R}^{k_n}$ in a neighborhood of β^*, where β^* exists and solves $S_n^*(\beta^*) = 0$. In addition, suppose that, for $n > N^*$:*

(i) *there exists a constant $0 < c < \infty$ such that $\|\mathscr{I}_n^{-1}(\beta^*)\|_\infty \le c$;*

(ii) *there exists $\varepsilon > 0$ that may depend only on c such that for all $\delta > 0$ there exists $N_\delta \ge N^*$ such that, for $n > N_\delta$,*

$$\Pr\left\{\sup_{\|\beta-\beta^*\|_\infty<\varepsilon} \|H_n(\beta) - \mathscr{I}_n(\beta^*)\|_\infty > \frac{1}{2c}\right\} < \delta;$$

(iii) $\|S_n(\beta^*)\|_\infty \to_P 0$.

Then, as $n \to \infty$, a unique solution $\{\beta\colon S_n(\hat{\beta}) = 0\}$ exists in a neighborhood about β^* with probability going to 1, and $\|\hat{\beta} - \beta^*\|_\infty = O_p(\|S_n(\beta^*)\|_\infty)$.

Since the existence of β^* has already been demonstrated for sufficiently large n, and $\omega_n(\cdot, \beta)$ is continuous in β, the proof of part (b) follows from Lemma 1 if the validity of conditions (i) to (iii) can be established.

To prove (i), we need to show that $\|\mathscr{I}_n^{-1}(\beta^*)\|_\infty$ remains bounded as $n \to \infty$. The triangle inequality yields that

$$\|\mathscr{I}_n^{-1}(\beta^*)\|_\infty \le \|\mathscr{I}_n^{-1}(\beta^*) - \mathscr{I}_n^{-1}(\beta^{**})\|_\infty + \|\mathscr{I}_n^{-1}(\beta^{**})\|_\infty, \tag{14}$$

where $\mathscr{I}_n(\beta^{**})$ is given in (10). From (11), the latter term is bounded by M_3. From an inequality arising out of matrix perturbation theory [Golub and Van Loan (1989)],

$$\begin{aligned} &\|\mathscr{I}_n^{-1}(\beta^*) - \mathscr{I}_n^{-1}(\beta^{**})\|_\infty \\ &\quad \le \frac{\|\mathscr{I}_n(\beta^*) - \mathscr{I}_n(\beta^{**})\|_\infty \|\mathscr{I}_n^{-1}(\beta^{**})\|_\infty^2}{1 - \|\mathscr{I}_n^{-1}(\beta^{**})\|_\infty \|\mathscr{I}_n(\beta^*) - \mathscr{I}_n(\beta^{**})\|_\infty}, \end{aligned} \tag{15}$$

whenever $\|\mathscr{I}_n^{-1}(\beta^{**})\|_\infty \|\mathscr{I}_n(\beta^*) - \mathscr{I}_n(\beta^{**})\|_\infty \le 1$. Using (6), it follows easily from (11), (12) and Lemma A.1 that there exists an $N_3 < \infty$ such that $\|\mathscr{I}_n^{-1}(\beta^{**})\|_\infty \|\mathscr{I}_n(\beta^*) - \mathscr{I}_n(\beta^{**})\|_\infty$ is less than 1 for $n > N_3$, and therefore that $\|\mathscr{I}_n^{-1}(\beta^*)\|_\infty \le M_7$, where $M_7 \le M_3 + 1 < \infty$. Without loss of generality, we may set N^* (described in the statement of Lemma 1) to $\max\{N_3, N_\varepsilon\}$, where N_ε was determined in Section 3.1.

Condition (ii) of Lemma 1 will follow if there exists an $\varepsilon > 0$ such that

$$\Pr\left\{\sup_{\|\beta-\beta^*\|_\infty<\varepsilon} \|H_n(\beta) - \mathscr{I}_n(\beta^*)\|_\infty > \frac{1}{2M_7}\right\} \tag{16}$$

goes to 0 as $n \to \infty$. By the triangle inequality,

$$\|H_n(\beta) - \mathscr{I}_n(\beta^*)\|_\infty \le \|H_n(\beta^*) - \mathscr{I}_n(\beta^*)\|_\infty + \|H_n(\beta) - H_n(\beta^*)\|_\infty,$$

and the desired result will follow if there exists an $\varepsilon > 0$ such that, as $n \to \infty$,

$$\Pr\left\{\|H_n(\beta^*) - \mathscr{I}_n(\beta^*)\|_\infty > \frac{1}{4M_7}\right\} + \Pr\left\{\sup_{\|\beta-\beta^*\|_\infty < \varepsilon} \|H_n(\beta) - H_n(\beta^*)\|_\infty > \frac{1}{4M_7}\right\} \to 0.$$

To see that the first term goes to 0, we may begin by writing

$$\|H_n(\beta^*) - \mathscr{I}_n(\beta^*)\|_\infty = \left\|\frac{k_n}{n}\sum_{i=1}^{n}(A_i - E[A_i])\right\|_\infty,$$

where the matrix $A_i = (A_{igg'})$ and

$$A_{igg'} = \int_0^1 f_g(Z_i(u)) f_{g'}(Z_i(u)) \omega_n(Z_i(u), \beta^*) Y_i(u)\, du.$$

The properties of B-splines and (13) immediately yield that $\Pr\{|A_{igg'} - E[A_{igg'}]| \le M_8\} = 1$ for $M_8 < \infty$. In addition,

$$\begin{aligned}
\operatorname{Var}\left[\sum_{i=1}^{n} A_{igg'}\right] &\le \sum_{i=1}^{n} E\left[A_{igg'}^2\right] \\
&\le M_6^2 \sum_{i=1}^{n} E\left[\int_0^1 f_g^2(Z_i(u)) f_{g'}^2(Z_i(u)) Y_i(u)\, du\right] \\
&\le nM_6^2 \int_0^1 \int_\Theta f_g^2(\theta) f_{g'}^2(\theta) p(u,\theta) h_Q(\theta; u)\, d\theta\, du \\
&\le nM_1 M_6^2 \int_0^1 \int_\Theta f_g^2(\theta)\, d\theta\, du,
\end{aligned}$$

where the last inequality follows from the facts that (i) the marginal density of $Z(\cdot)$ [i.e., $h_Q(\cdot;\cdot)$] is bounded by M_1, and (ii) $p(u,\theta)$ and $f_{g'}^2(\cdot)$ are each bounded above by 1. Evaluation of the integrals in the last term shows that $\operatorname{Var}[\sum_{i=1}^n A_{igg'}] \le M_1 M_6^2 n\overline{\Delta}_n$. Applying Bernstein's inequality [Serfling (1980)],

$$\Pr\left\{\frac{k_n}{n}\left|\sum_{i=1}^{n}\left(A_{igg'} - E[A_{igg'}]\right)\right| > \frac{1}{4M_7}\right\} \le 2\exp\left\{-n\overline{\Delta}_n M_9\right\}$$

for some constant $M_9 < \infty$ that depends on M_1, M_6 and M_8. This yields that

$$\Pr\left\{\|H_n(\beta^*) - \mathscr{I}_n(\beta^*)\|_\infty > \frac{1}{4M_7}\right\} \le 2k_n^2 \exp\left\{-n\overline{\Delta}_n M_9\right\}, \tag{17}$$

which goes to 0 as $n \to \infty$ since $n\overline{\Delta}_n \to \infty$.

Now, to demonstrate (16), we only need to show that

$$\Pr\left\{\sup_{\|\beta-\beta^*\|_\infty < \varepsilon} \|H_n(\beta) - H_n(\beta^*)\|_\infty > \frac{1}{4M_7}\right\}$$

goes to 0 as $n \to \infty$ for some $\varepsilon > 0$. Properties of sup-norm immediately give that

$$\|H_n(\beta) - H_n(\beta^*)\|_\infty \le \left\| \frac{k_n}{n} \sum_{i=1}^{n} \int_0^1 \mathbf{f}(Z_i(u))\mathbf{f}'(Z_i(u))Y_i(u)\,du \right\|_\infty \times \|\omega_n(\cdot, \beta) - \omega_n(\cdot, \beta^*)\|_\infty.$$

By the triangle inequality, the first term on the right-hand side may be bounded by

$$\left\| \frac{k_n}{n} \sum_{i=1}^{n} (B_i - E[B_i]) \right\|_\infty + \left\| \frac{k_n}{n} \sum_{i=1}^{n} E[B_i] \right\|_\infty,$$

where the matrix $B_i = (B_{igg'})$ has elements

$$B_{igg'} = \int_0^1 f_g(Z_i(u)) f_{g'}(Z_i(u)) Y_i(u)\,du.$$

Since $\|\omega_n(\cdot, \beta^*)\|_\infty \le M_6$ and $\omega_n(\cdot, \beta)$ is continuous and differentiable in β, applying the mean value theorem yields

$$\sup_{\|\beta - \beta^*\|_\infty < \varepsilon} \|\omega_n(\cdot, \beta) - \omega_n(\cdot, \beta^*)\|_\infty \le M_{10}\,\varepsilon$$

for some constant $M_{10} < \infty$.

Thus,

$$\Pr\left\{ \sup_{\|\beta - \beta^*\|_\infty < \varepsilon} \|H_n(\beta) - H_n(\beta^*)\|_\infty > \frac{1}{4M_7} \right\}$$
$$\le \Pr\left\{ \left\| \frac{k_n}{n} \sum_{i=1}^{n} (B_i - E[B_i]) \right\|_\infty > \frac{1}{8M_7 M_{10}\,\varepsilon} \right\}$$
$$+ \Pr\left\{ \left\| \frac{k_n}{n} \sum_{i=1}^{n} E[B_i] \right\|_\infty > \frac{1}{8M_7 M_{10}\,\varepsilon} \right\}.$$

Now, by Lemma A.1, we may find a constant M_{11} such that $\|(k_n/n)\sum_{i=1}^n E[B_i]\|_\infty \le M_{11}$, and hence the latter probability can be made equal to 0 for any $n > N^*$ by choosing $\varepsilon < (8M_7 M_{10} M_{11})^{-1}$. Using similar arguments to those before, the first term on the right-hand side goes to 0 as $n \to \infty$ after applying Bernstein's inequality. Thus, an $\varepsilon > 0$ exists independently of $n > N^*$ such that condition (ii) of Lemma 1 holds.

Finally, we must show that $\|S_n(\beta^*)\|_\infty \to_P 0$. From (3), the score function $S_n(\beta^*)$ may be expressed as the sum of two pieces. The first piece, which does not involve β^*, will be denoted as S_n^M, and is a stochastic integral with respect to a martingale process. The second piece, which does involve β^*, will be denoted as $S_n^{NM}(\beta^*)$. It is therefore sufficient to establish that

$$\|S_n^M\|_\infty = \left\| \frac{k_n}{n} \sum_{i=1}^n \int_0^1 \mathbf{f}(Z_i(u))\, dM_i(u) \right\|_\infty \to_P 0 \tag{18}$$

and

$$\begin{aligned} &\|S_n^{NM}(\beta^*)\|_\infty \\ &\quad = \left\| \frac{k_n}{n} \sum_{i=1}^n \int_0^1 \mathbf{f}(Z_i(u))[\omega(Z_i(u)) - \omega_n(Z_i(u), \beta^*)] Y_i(u)\, du \right\|_\infty \\ &\quad \to_P 0. \end{aligned} \tag{19}$$

Since each element of the vector $\sum_{i=1}^n \int_0^1 \mathbf{f}(Z_i(u))\, dM_i(u)$ is a sum of bounded, mean zero random variables with variance bounded by a term of order $n\overline{\Delta}_n$, Bernstein's inequality may be used to show that $\|S_n^M\|_\infty \to_P 0$. For the latter term,

$$\begin{aligned} \|S_n^{NM}(\beta^*)\|_\infty &\le \left\| \frac{k_n}{n} \sum_{i=1}^n \int_0^1 \mathbf{f}(Z_i(u)) Y_i(u)\, du \right\|_\infty \|\omega - \omega_n(\cdot, \beta^*)\|_\infty \\ &= \left(\left\| \frac{k_n}{n} \sum_{i=1}^n (D_i - E[D_i]) \right\|_\infty + \left\| \frac{k_n}{n} \sum_{i=1}^n E[D_i] \right\|_\infty \right) \\ &\qquad \times \|\omega_n(\cdot, \beta^*) - \omega\|_\infty, \end{aligned}$$

where $D_i = \int_0^1 \mathbf{f}(Z_i(u)) Y_i(u)\, du$.

Using arguments similar to those in the previous section, the term in the parentheses may be bounded in probability by applying Bernstein's inequality and Lemma A.1; from (8) and (12), $\|\omega_n(\cdot, \beta^*) - \omega\|_\infty = O(\overline{\Delta}_n^2)$ and hence $\|S_n^{NM}(\beta^*)\|_\infty = O_p(\overline{\Delta}_n^2)$.

Thus, the conditions of Lemma 1 have been demonstrated to hold for $S_n(\beta)$ with probability going to 1. Therefore, $\hat{\beta}$ exists with probability going to 1 and satisfies $\|\hat{\beta} - \beta^*\|_\infty \to_P 0$. To complete the proof, we must show that $\|\omega_n(\cdot, \hat{\beta}) - \omega\|_\infty \to_P 0$. However, this follows immediately from the triangle inequality, continuity and the fact that $\|\omega_n(\cdot, \beta^*) - \omega\|_\infty = O(\overline{\Delta}_n^2)$. $\square$

4. Asymptotic normality. For $\nu \in \Theta^o$ (the interior of Θ), we desire to prove that $\omega_n(\nu, \hat{\beta}) - \omega(\nu)$, properly normalized, converges pointwise in law to a normal random variable. Simple manipulations bring us to the following sufficient result, stated below as a theorem.

THEOREM 2. *For $\nu \in \Theta^o$ and $k_n = n^\phi$, $\phi \in (\frac{1}{5}, 1)$,*

$$\left(\frac{n}{k_n} \right)^{1/2} \|\omega_n(\cdot, \beta^*) - \omega\|_\infty \to 0 \tag{20}$$

and

$$\left(\frac{n}{k_n}\right)^{1/2} \frac{\log \omega_n(\nu, \hat{\beta}) - \log \omega_n(\nu, \beta^*)}{\sigma(\nu, \pi_n)} \to_L N(0,1), \tag{21}$$

where $\sigma^2(\nu, \pi_n)$ *is such that the variance of* (21) *converges to unity and* $\pi_n = (\nu - \tau_{g\nu})/\overline{\Delta}_n$ *for* $g_\nu = \{r: \tau_r \le \nu < \tau_{r+1}\}$.

In the proof of the theorem, we shall demonstrate that $\sigma^2(\nu, \pi_n)$ is the variance of

$$\left(\frac{n}{k_n}\right)^{1/2} (S_n^M)' \mathscr{I}_n^{-1}(\beta^*)\mathbf{f}(\nu),$$

where S_n^M is defined in (18). The above is a sum of stochastic integrals, the ith being taken with respect to the $\mathscr{F}_t$-martingale $M_i(t)$, $i = 1, \dots, n$. The explicit form of the variance $\sigma^2(\nu, \pi_n)$ is given in Lemma A.5, and a consistent estimate of it is provided at the end of this paper. For a fixed sample size n, this variance depends on ν as well as its relative position between the bracketing knots (i.e., π_n). This is not immediately obvious nor intuitively appealing; however, it is relatively easy to explain, especially for linear B-splines. For large n, we demonstrate this phenomenon for the case of linear B-splines in Appendix B and briefly discuss its implications. The derivation for higher-order splines proceeds similarly.

PROOF OF THEOREM 2

Proof of (20). It was proved in Section 3.1 that $\|\omega_n(\cdot, \beta^*) - \omega\|_\infty = O(\overline{\Delta}_n^2)$; thus,

$$\left(\frac{n}{k_n}\right)^{1/2} \|\omega_n(\cdot, \beta^*) - \omega\|_\infty = \left(\frac{n}{k_n}\right)^{1/2} O(\overline{\Delta}_n^2) = O\left(\left(\frac{n}{k_n^5}\right)^{1/2}\right).$$

Since $k_n = n^\phi$, $\phi \in (\frac{1}{5}, 1)$, this goes to 0 as $n \to \infty$, proving (20). □

Proof of (21). A first-order Taylor series expansion of $S_n(\hat{\beta})$ around β^* yields that

$$S_n(\hat{\beta}) = S_n(\beta^*) - H_n(\tilde{\beta})(\hat{\beta} - \beta^*),$$

where $\tilde{\beta}$ lies on a line segment connecting $\hat{\beta}$ and β^*. For reference, the definitions of $S_n(\beta)$ and $H_n(\beta)$ are given by (3) and (5), respectively. Using the fact that $S_n(\hat{\beta}) = 0$, then after some algebra one obtains the expression

$$\left(\frac{n}{k_n}\right)^{1/2} (\hat{\beta} - \beta^*)'\mathbf{f}(\nu) = \left(\frac{n}{k_n}\right)^{1/2} (S_n(\beta^*))' H_n^{-1}(\tilde{\beta})\mathbf{f}(\nu). \tag{22}$$

Note that the left-hand side of this expression is the numerator of (21). Now, from (3), the score function $S_n(\beta^*)$ may be expressed as $S_n^M + S_n^{NM}(\beta^*)$;

definitions of the respective terms may be obtained from (18) and (19). Returning to (22), we see that

$$\begin{aligned}\left(\frac{n}{k_n}\right)^{1/2} &(\hat{\beta}-\beta^*)'\mathbf{f}(\nu)\\ &=\left(\frac{n}{k_n}\right)^{1/2}\left[S_n^M+S_n^{NM}(\beta^*)\right]'H_n^{-1}(\tilde{\beta})\mathbf{f}(\nu)\\ &=\left(\frac{n}{k_n}\right)^{1/2}\left[(S_n^M)'H_n^{-1}(\tilde{\beta})\mathbf{f}(\nu)+(S_n^{NM}(\beta^*))'H_n^{-1}(\tilde{\beta})\mathbf{f}(\nu)\right].\end{aligned}$$

By adding and subtracting $\mathscr{I}_n^{-1}(\beta^*)$ from $H_n^{-1}(\tilde{\beta})$ and rearranging terms, the proof of (21) now follows if as $n\to\infty$,

$$\left(\frac{n}{k_n}\right)^{1/2}\frac{(S_n^M)'\mathscr{I}_n^{-1}(\beta^*)\mathbf{f}(\nu)}{\sigma(\nu,\pi_n)}\to_L N(0,1), \tag{23}$$

$$\left(\frac{n}{k_n}\right)^{1/2}\left\|(S_n^M)'\left[H_n^{-1}(\tilde{\beta})-\mathscr{I}_n^{-1}(\beta^*)\right]\mathbf{f}\right\|_\infty\to_P 0, \tag{24}$$

$$\left(\frac{n}{k_n}\right)^{1/2}\left\|(S_n^{NM}(\beta^*))'\mathscr{I}_n^{-1}(\beta^*)\mathbf{f}\right\|_\infty\to_P 0 \tag{25}$$

and

$$\left(\frac{n}{k_n}\right)^{1/2}\left\|(S_n^{NM}(\beta^*))'\left[H_n^{-1}(\tilde{\beta})-\mathscr{I}_n^{-1}(\beta^*)\right]\mathbf{f}\right\|_\infty\to_P 0, \tag{26}$$

where the scalar $\sigma^2(\nu,\pi_n)=\mathrm{Var}((n/k_n)^{1/2}(S_n^M)'\mathscr{I}_n^{-1}(\beta^*)\mathbf{f}(\nu))$. The proofs of (24)–(26) are done in Lemmas A.2–A.4. It is shown in Lemma A.5 that $\sigma^2(\nu,\pi_n)$ is bounded away from 0 and ∞. Hence, the proof of (23), done below, completes the proof of (21) and therefore Theorem 2.

Proof of (23). Since $\mathscr{I}_n^{-1}(\beta^*)$ and $\sigma^2(\nu,\pi_n)$ are deterministic, we may write each component of (23) as a stochastic integral of a predictable process with respect to a martingale process [Fleming and Harrington (1991)]; that is,

$$\left(\frac{k_n}{n}\right)^{1/2}\sum_{i=1}^n\int_0^1\frac{\mathbf{f}'(Z_i(u))\mathscr{I}_n^{-1}(\beta^*)\mathbf{f}(\nu)}{\sigma(\nu,\pi_n)}\,dM_i(u). \tag{27}$$

This converges in law to a standard normal random variable by the martingale central limit theorem provided that the following two sufficient conditions are met [Andersen and Gill (1982)]:

(I) As $n\to\infty$,

$$\frac{k_n}{n}\sum_{i=1}^n\int_0^1\left[\frac{\mathbf{f}'(Z_i(u))\mathscr{I}_n^{-1}(\beta^*)\mathbf{f}(\nu)}{\sigma(\nu,\pi_n)}\right]^2\omega(Z_i(u))Y_i(u)\,du\to_P 1.$$

(II) As $n \to \infty$ and for all $\varepsilon > 0$,

$$\frac{k_n}{n}\sum_{i=1}^{n}\int_0^1\left[\frac{\mathbf{f}'(Z_i(u))\mathscr{I}_n^{-1}(\beta^*)\mathbf{f}(\nu)}{\sigma(\nu,\pi_n)}\right]^2 I_\varepsilon(Z_i(u),\nu)\,\omega(Z_i(u))Y_i(u)\,du \to_P 0$$

where

$$I_\varepsilon(Z(u),\nu) = I\left\{\left(\frac{k_n}{n}\right)^{1/2}\left|\frac{\mathbf{f}'(Z(u))\mathscr{I}_n^{-1}(\beta^*)\mathbf{f}(\nu)}{\sigma(\nu,\pi_n)}\right| > \varepsilon\right\}.$$

Result (I) immediately follows from the definition of $\sigma^2(\nu, \pi_n)$, Lemma A.5 and the weak law of large numbers for triangular arrays [cf. Feller (1971)].

To prove (II), note that

$$\sup_{(\nu,\theta)\in\Theta^2}\left|\mathbf{f}'(\theta)\mathscr{I}_n^{-1}(\beta^*)\mathbf{f}(\nu)\right| = \sup_{(\nu,\theta)\in\Theta^2}\left|\left(\mathscr{I}_n^{-1}(\beta^*)\mathbf{f}(\theta)\right)'\mathbf{f}(\nu)\right| \le \left\|\mathscr{I}_n^{-1}(\beta^*)\right\|_\infty,$$

which is bounded by a finite constant for large n; see Section 3.2 for details. Thus, for any $\varepsilon > 0$, there exists N_ε such that for all $n > N_\varepsilon$, $I_\varepsilon(Z(u), \nu) = 0$, proving (II). Hence, the martingale central limit theorem [cf. Fleming and Harrington (1991), Theorem 5.3.4] may now be applied to prove that (27) converges in law to a $N(0, 1)$ random variable, completing the proof of Theorem 2. □

5. Remarks. These results have been obtained assuming $\omega(\cdot)$ is twice-continuously differentiable. From the proof in Section 4, we see that the necessary convergence rates depend on this assumption only through (8) and (12). Thus, if we had instead assumed that $\omega(\cdot)$ was p-times continuously differentiable, the rates for (8) and (12) would be on the order of $\overline{\Delta}_n^k$ for a kth-order spline, where $k \le p$. The asymptotic normality results then apply for $k_n = n^\phi$ when $\phi > 1/(2k + 1)$.

In the course of proving Theorem 2, a stronger result than pointwise asymptotic normality has actually been demonstrated. Specifically, note that (27) is composed of a sum of mean zero random variables, each of which is bounded in absolute value by a term that is $O(k_n/n)$. In addition, the variance of the sum is bounded above and below by terms of $O(k_n/n)$. These facts, plus a corollary to the central limit theorem [cf. Chung (1974), page 201], yield convergence in distribution *uniformly* for $\nu \in \Theta^o$. Similar results were proved by Stone (1990, 1991).

Since this paper was submitted, an important related work of Stone (1994) has been brought to our attention. Stone (1994) elegantly established optimal L_2 rates of convergence for a wide class of spline-based estimators in regression, generalized regression (e.g., GLM's) and density estimation in a unified framework. Kooperberg, Stone and Truong (1995b) extend those results to the case of hazard regression for time-independent covariates. In both papers, the optimal L_2 rate of convergence for functions satisfying a similar smoothness condition to $\omega(\cdot)$ corresponds to $k_n = n^{1/5}$. We have not specifically addressed this question here. However, from the proof in Section 4, it is clear that (21) converges to $Z + W$ for $Z \sim N(0,1)$ and some random variable W. The random variable W is $O_p(1)$ at $k_n = n^{1/5}$, and respectively satisfies $P\{W = 0\} = 1$ or $P\{W = \infty\} = 1$ according to whether $k_n = n^{1/5+\varepsilon}$ or $k_n = n^{1/5-\varepsilon}$ for some $\varepsilon > 0$. Given this usual bias–variance trade-off and the results of Stone (1994) and Kooperberg, Stone and Truong (1995b), we conjecture (but have not proved) that $n^{1/5}$ is also the optimal L_2 rate of convergence for this problem. It may be possible to extend the results of Kooperberg, Stone and Truong (1995b) to establish this result directly.

Assuming that $n^{1/5}$ is, in fact, the optimal L_2 rate of convergence, a referee rightly points out that effective use of the asymptotic normality result (e.g., for constructing pointwise confidence intervals) requires undersmoothing the data. In other words, "mean zero" asymptotic normality occurs only for $k_n = n^{1/5+\varepsilon}$ for $\varepsilon > 0$, which implies that one must use fewer observations per interval than is needed to achieve the smallest possible mean squared error.

An interesting issue that arises out of the proof of asymptotic normality concerns the behavior of the variance. In particular, the variance of the MLE for large but fixed values of n is a polynomial function of the relative location between two adjacent knots. This can be seen in Appendix B, where such an expression is derived for linear B-splines. The variance in this case is minimized when the point ν is exactly halfway between the two bracketing knots, and grows to nearly three times larger at the endpoints. Although simple mathematical representations for the variance are not readily available in the general case, numerical calculations indicate that for even-order splines, the variance is indeed minimized halfway between the bracketing knots. Interestingly, similar calculations indicate that the opposite is true for odd-order splines. For example, the variance for a third-order spline (a piecewise quadratic polynomial) appears to be minimized at the knots, and attains a maximum at the midpoint. In practice, the variance $\sigma^2(\nu, \pi_n)$ cannot be calculated, and a consistent estimator must be used. Using condition (I) from the proof of Theorem 2 and plugging in consistent estimators for the unknowns, we obtain

$$\hat{\sigma}^2(\nu, \pi_n) = \frac{k_n}{n} \sum_{i=1}^{n} \int_0^1 \left[\mathbf{f}'(Z(u)) H_n^{-1}(\hat{\beta})\mathbf{f}(\nu)\right]^2 \omega_n\left(Z_i(u), \hat{\beta}\right) Y_i(u)\, du.$$

The results derived here are not particularly useful with regard to determining the optimal number or position of the knots to be used in any finite-dimensional problem. If, however, the interest lies in a specific set of points, it appears that both the spline order and (uniformly spaced) knots can be chosen to minimize the variance of the estimates at those particular points. In other cases, one might use some data-driven technique (e.g., Bayesian information criterion, cross-validation, etc.) to address the problem of knot selection in the finite-sample problem. However, determining the optimal number and position of these knots (as measured by some minimization criterion) is a much more difficult problem. Important contributions in this area have been made by Charles Kooperberg and his colleagues. For algorithms relevant to the hazard regression problem, see the recent work of Kooperberg, Stone and Truong (1995a).

APPENDIX A

Lemmas for Sections 3 and 4.

LEMMA A.1. *Let* $B = \lim_{n\to\infty} B_n$, *where the matrix*

$$B_n = k_n \int_0^1 \left[\int_\Theta \mathbf{f}(\theta)\mathbf{f}'(\theta) y_n(u, \theta)\, d\theta \right] du$$

and $y_n(u, \theta)$ *is a positive, continuous and deterministic function on* $[0, 1] \times \Theta$ *that is bounded away from* 0 *and* ∞ *for* $n \geq N$ *for some* $N < \infty$. *Then,* $\|B\|_\infty < \infty$ *and* $\|B^{-1}\|_\infty < \infty$.

PROOF. For large and fixed n, the mean value theorem for integrals [see Apostol (1957)] yields

$$(B_n)_{gg'} \approx k_n \int_0^1 y_n(u, \xi_g) \left[\int_\Theta f_g(\theta) f_{g'}(\theta)\, d\theta \right] du,$$

where $\xi_g \in \Theta$. Since the knots are equally spaced with mesh size $\overline{\Delta}_n$,

$$\int_\Theta f_g(\theta) f_{g'}(\theta)\, d\theta = \overline{\Delta}_n I^m_{|g-g'|},$$

where the I^m_j's are bounded nonnegative constants that depend only on the order of the spline m ($= p + 1$). In addition, they are positive if $|g - g'| \leq p$ and satisfy $I^m_0 + 2\sum_{j\geq 1} I^m_j = 1$ [Schumaker (1981)]. For example, if $\mathbf{f}(\cdot)$ is a linear B-spline basis, then

$$I^2_{|g-g'|} = \begin{cases} \frac{4}{6}, & \text{if } g = g', \\ \frac{1}{6}, & \text{if } |g - g'| = 1, \\ 0, & \text{if } |g - g'| > 1. \end{cases} \qquad (28)$$

Thus, for large n, $B_n \approx k_n \overline{\Delta}_n T_n D_n$, where $T_n = (t_{nij})$ is a $2p + 1$-banded symmetric $k_n \times k_n$ Toeplitz matrix such that $t_{nij} = I^m_{|i-j|}$ and $\|T_n\|_\infty = 1$ for each n, and D_n is a diagonal matrix with elements $d_{gg} = \int_0^1 y_n(u, \xi_g)\, du$.

The behavior of Toeplitz matrices and their inverses has been extensively studied. The eigenvalues of a general Toeplitz matrix can be shown to be asymptotically equivalent to those of a circulant matrix [Brillinger (1980), page 73]. In particular, the maximal eigenvalue, say $\lambda_{\max}$, of the matrix T_n converges asymptotically to unity, and the minimal eigenvalue converges to

$$\lambda_{\min} = \sum_{j=-p}^{p} I^m_{|j|} \cos\left(\frac{2\pi j}{p + p \bmod 2} \right).$$

Since $m = p + 1 < \infty$, it follows that $\lambda_{\min} > 0$ and thus T_n remains positive definite as $n \to \infty$. The statement of the theorem implies that there exist constants d_1 and d_2 independently of k_n for $n \geq N$ such that $0 < d_1 < d_{gg} < d_2 < \infty$ for $g = 1, \dots, k_n$. Thus, D_n is positive definite as $n \to \infty$, and $\|D_n\|_\infty < d_2$. Finally, since $\lim_{n\to\infty} k_n \bar{\Delta}_n = 1$, it follows that the matrix $B = \lim_{n\to\infty} B_n$ is positive definite and thus invertible, with $\|B\|_\infty \leq 2d_2 < \infty$.

To determine an upper bound on $\|B^{-1}\|_\infty$, it is helpful to first recall the facts that for a nonnegative definite symmetric matrix M, $\|M\|_2$ is equal to the largest eigenvalue of M and is bounded above by $\|M\|_\infty$ [cf. Golub and Van Loan (1989), page 58]. Thus, it follows from earlier results that $\|B^{-1}\|_2 < (d_1\lambda_{\min})^{-1} < \infty$. Now, since $\|B\|_\infty$ is bounded by some finite constant K, the results of Demko [(1977), Theorem 2.2] immediately yield the existence of constants $K' < \infty$ and $r \in (0, 1)$ depending only on $\lambda_{\min}$, d_1, d_2 and the spline order m such that $|(B^{-1})_{ij}| \leq K' r^{|i-j|}$ and $\|B^{-1}\|_\infty \leq 2K'(1 - r)$. Hence, not only is B^{-1} bounded in sup-norm, but its elements decay exponentially fast to 0 as one moves away from the main diagonal. □

LEMMA A.2. *As* $n \to \infty$,

$$\left(\frac{n}{k_n}\right)^{1/2} \left\|(S_n^M)'\left[H_n^{-1}(\tilde{\beta}) - \mathscr{I}_n^{-1}(\beta^*)\right]\mathbf{f}\right\|_\infty \to_P 0.$$

PROOF. Let $\alpha \in [0, \frac{1}{2}]$ be a constant. We may write

$$\begin{aligned}
&\left(\frac{n}{k_n}\right)^{1/2} \left\|(S_n^M)'\left[H_n^{-1}(\tilde{\beta}) - \mathscr{I}_n^{-1}(\beta^*)\right]\mathbf{f}\right\|_\infty \\
&\qquad \leq \left(\frac{n}{k_n}\right)^{1/2} \|S_n^M\|_\infty \left\|H_n^{-1}(\tilde{\beta}) - \mathscr{I}_n^{-1}(\beta^*)\right\|_\infty \\
&\qquad = \left(\frac{n}{k_n}\right)^{(1/2)-\alpha} \|S_n^M\|_\infty \left(\frac{n}{k_n}\right)^{\alpha} \left\|H_n^{-1}(\tilde{\beta}) - \mathscr{I}_n^{-1}(\beta^*)\right\|_\infty.
\end{aligned}$$

Referring to Section 3.2, $\|S_n^M\|_\infty \to_P 0$ by Bernstein's inequality. In particular, for all $\delta > 0$,

$$\Pr\left\{\|S_n^M\|_\infty > \delta\right\} \leq 2k_n \exp\left\{-O\left(\frac{n}{k_n}\right)\right\},$$

which goes to 0 as $n \to \infty$. A similar argument therefore yields that

$$\Pr\left\{\left(\frac{n}{k_n}\right)^{(1/2)-\alpha} \|S_n^M\|_\infty > \delta\right\} \leq 2k_n \exp\left\{-O\left(\left(\frac{n}{k_n}\right)^{\alpha+1/2}\right)\right\},$$

which still goes to 0 as $n \to \infty$.

Since $\|\mathscr{I}_n^{-1}(\beta^*)\|_\infty$ remains bounded for large n, arguments similar to those in Section 3.2 yield that

$$\left(\frac{n}{k_n}\right)^\alpha \left\| H_n^{-1}(\tilde{\beta}) - \mathscr{I}_n^{-1}(\beta^*)\right\|_\infty \to_P 0$$

if it can be demonstrated that

$$\left(\frac{n}{k_n}\right)^\alpha \left\| H_n(\tilde{\beta}) - \mathscr{I}_n(\beta^*)\right\|_\infty \to_P 0.$$

By the triangle inequality, this will follow if

$$\left(\frac{n}{k_n}\right)^\alpha \left\| H_n(\beta^*) - \mathscr{I}_n(\beta^*)\right\|_\infty \to_P 0$$

and

$$\left(\frac{n}{k_n}\right)^\alpha \left\| H_n(\tilde{\beta}) - H_n(\beta^*)\right\|_\infty \to_P 0.$$

Using arguments similar to those used to obtain (16), the first follows from Bernstein's inequality. For the latter, the continuity of $H_n(\beta)$, the mean value theorem and the fact that $\tilde{\beta}$ lies on a line segment between $\hat{\beta}$ and β^* imply that

$$\begin{aligned}
\left(\frac{n}{k_n}\right)^\alpha \left\| H_n(\tilde{\beta}) - H_n(\beta^*)\right\|_\infty &\le \left(\frac{n}{k_n}\right)^\alpha M_{12}\|\tilde{\beta} - \beta^*\|_\infty \\
&\le \left(\frac{n}{k_n}\right)^\alpha M_{12}\|\hat{\beta} - \beta^*\|_\infty \\
&= \left(\frac{n}{k_n}\right)^\alpha M_{13}\|S_n(\beta^*)\|_\infty \\
&\le \left(\frac{n}{k_n}\right)^\alpha M_{13}\left(\|S_n^M\|_\infty + \left\|S_n^{NM}(\beta^*)\right\|_\infty\right)
\end{aligned}$$

for appropriately chosen constants $M_{12} < \infty$ and $M_{13} < \infty$. It was just shown that $(n/k_n)^\alpha\|S_n^M\|_\infty \to_P 0$ for any $\alpha \in [0, \frac{1}{2}]$. Hence, since

$$\left(\frac{n}{k_n}\right)^\alpha \left\|S_n^{NM}(\beta^*)\right\|_\infty = \left(\frac{n}{k_n}\right)^\alpha O_p\left(\overline{\Delta}_n^2\right),$$

the latter term also converges to 0 in probability for any $k_n = n^\phi$, $\phi \in (\frac{1}{5}, 1)$ as long as $\alpha \le 0.5$. □

LEMMA A.3. *As* $n \to \infty$,

$$\left(\frac{n}{k_n}\right)^{1/2} \left\|\left(S_n^{NM}(\beta^*)\right)'\mathscr{I}_n^{-1}(\beta^*)\mathbf{f}\right\|_\infty \to_P 0.$$

PROOF. Note that

$$\left(\frac{n}{k_n}\right)^{1/2}\left\|\left(S_n^{NM}(\beta^*)\right)'\mathscr{I}_n^{-1}(\beta^*)\mathbf{f}\right\|_\infty \le \left(\frac{n}{k_n}\right)^{1/2}\left\|S_n^{NM}(\beta^*)\right\|_\infty\left\|\mathscr{I}_n^{-1}(\beta^*)\right\|_\infty.$$

Since $\|\mathscr{I}_n^{-1}(\beta^*)\|_\infty$ remains bounded as $n \to \infty$ (Section 3.2) and $S_n^{NM}(\beta^*) = O_p(\overline{\Delta}_n^2)$,

$$\begin{aligned}\left(\frac{n}{k_n}\right)^{1/2}\left\|S_n^{NM}(\beta^*)\right\|_\infty\left\|\mathscr{I}_n^{-1}(\beta^*)\right\|_\infty &\le \left(\frac{n}{k_n}\right)^{1/2} O_p\left(\overline{\Delta}_n^2\right)\left\|\mathscr{I}_n^{-1}(\beta^*)\right\|_\infty \\ &= O_p\left(\left(\frac{n}{k_n^5}\right)^{1/2}\right).\end{aligned}$$

Since $k_n = n^\phi$, $\phi \in (\frac{1}{5}, 1)$, this goes to 0 as $n \to \infty$. □

LEMMA A.4. *As* $n \to \infty$,

$$\left(\frac{n}{k_n}\right)^{1/2}\left\|\left(S_n^{NM}(\beta^*)\right)'\left[H_n^{-1}(\tilde{\beta}) - \mathscr{I}_n^{-1}(\beta^*)\right]\mathbf{f}\right\|_\infty \to_P 0.$$

PROOF. Properties of sup-norm imply that

$$\begin{aligned}&\left(\frac{n}{k_n}\right)^{1/2}\left\|\left(S_n^{NM}(\beta^*)\right)'\left[H_n^{-1}(\tilde{\beta}) - \mathscr{I}_n^{-1}(\beta^*)\right]\mathbf{f}\right\|_\infty \\ &\qquad \le \left(\frac{n}{k_n}\right)^{1/2}\left\|S_n^{NM}(\beta^*)\right\|_\infty\left\|H_n^{-1}(\tilde{\beta}) - \mathscr{I}_n^{-1}(\beta^*)\right\|_\infty.\end{aligned}$$

For $k_n = n^\phi$, $\phi \in (\frac{1}{5}, 1)$, we know from Lemmas A.2 and A.3 that $\|H_n^{-1}(\tilde{\beta}) - \mathscr{I}_n^{-1}(\beta^*)\|_\infty \to_P 0$ and $(n/k_n)^{1/2}\|S_n^{NM}(\beta^*)\|_\infty \to_P 0$; the desired result follows from Slutsky's theorem. □

LEMMA A.5. *For* $\nu \in \Theta^o$ *and as* $n \to \infty$, $\sigma^2(\nu, \pi_n)$ *is bounded away from* 0 *and* ∞.

PROOF. Using the results of Aalen (1978),

$$\begin{aligned}\sigma^2(\nu, \pi_n) &= E\left[\frac{k_n}{n}\sum_{i=1}^n \int_0^1 \left[\mathbf{f}'(Z_i(u))\mathscr{I}_n^{-1}(\beta^*)\mathbf{f}(\nu)\right]^2 \omega(Z_i(u))Y_i(u)\,du\right] \\ &= k_n\int_0^1 E\left[\left[\mathbf{f}'(Z(u))\mathscr{I}_n^{-1}(\beta^*)\mathbf{f}(\nu)\right]^2 \omega(Z(u))I\{X \ge u\}\,du\right] \\ &= k_n\int_0^1\int_\Theta \left[\mathbf{f}'(\theta)\mathscr{I}_n^{-1}(\beta^*)\mathbf{f}(\nu)\right]^2 \omega(\theta)p(u,\theta)h_Q(\theta; u)\,d\theta\,du.\end{aligned}$$

The behavior of the variance is primarily determined by the bilinear form $\mathbf{f}'(\theta)\mathscr{I}_n^{-1}(\beta^*)\mathbf{f}(\nu)$. However, note that for fixed ν, we may rewrite this bilinear form as $\gamma(\nu)'\mathbf{f}(\theta)$, which is an mth-order B-spline in θ with coefficients $\gamma(\nu) = \mathscr{I}_n^{-1}(\beta^*)\mathbf{f}(\nu)$. It follows under the assumptions of this paper that we may find constants $0 < w$, $W < \infty$ such that $wQ \le \sigma^2(\nu, \pi_n) \le QW$, where

$$\begin{aligned} Q &= k_n \int_\Theta [\gamma(\nu)'\mathbf{f}(\theta)]^2 \, d\theta \\ &= k_n \overline{\Delta}_n \gamma(\nu)' T_n \gamma(\nu), \end{aligned}$$

where T_n is the banded Toeplitz matrix defined in Lemma A.1. Since T_n is positive definite for all n and there is at least one (and at most m) of the $\gamma_g(\nu)$ that are nonzero for every $\nu \in \Theta^o$, the quadratic form $Q > 0$. Thus, $\sigma^2(\nu, \pi_n)$ is bounded away from 0 as $n \to \infty$. In addition,

$$\sup_{\nu \in \Theta^o} |Q| \le \left\| \mathscr{I}_n^{-1}(\beta^*) \right\|_\infty^2 \|T_n\|_\infty.$$

Since $\|T_n\|_\infty = 1$ and $\|\mathscr{I}_n^{-1}(\beta^*)\|_\infty$ is bounded, Q is also bounded above and therefore $\sigma^2(\nu, \pi_n)$ is bounded away from ∞. □

APPENDIX B

Derivation of variance in linear case. Let $\mathbf{f}(\cdot)$ denote the usual linear B-spline basis on a uniform mesh. Since our main interest lies in asymptotic results, we assume that the sample size n is large. Then, from the proof of Lemma A.5,

$$\begin{aligned} \sigma^2(\nu, \pi_n) &= k_n \int_0^1 \int_\Theta \left[\mathbf{f}'(\theta)\mathscr{I}_n^{-1}(\beta^*)\mathbf{f}(\nu)\right]^2 \omega(\theta) p(u, \theta) h_Q(\theta; u) \, d\theta \, du \\ &\approx k_n \overline{\Delta}_n \gamma(\nu)' T_n \gamma(\nu) \int_0^1 \omega(\nu) p(u, \nu) h_Q(\nu; u) \, du, \end{aligned}$$

where $\gamma(\nu) = \mathscr{I}_n^{-1}(\beta^*)\mathbf{f}(\nu)$. The last step follows from the continuity of the integrand and the fact that $\gamma(\nu)'\mathbf{f}(\theta)$ has significant mass only in a very small neighborhood about the point $\theta = \nu$ for large n. If we let $\alpha(\nu) = \int_0^1 \omega(\nu) p(u, \nu) h_Q(\nu; u) \, du$, and note that

$$\mathscr{I}_n^{-1}(\beta^*) \approx \frac{1}{k_n \overline{\Delta}_n} D_n^{-1} T_n^{-1},$$

where D_n is a diagonal matrix having elements $\alpha(\xi_g)$ for interior knots $\xi_g \in \Theta^o$, then we may write

$$\sigma^2(\nu, \pi_n) \approx \alpha(\nu)\left(D_n^{-1}\mathbf{h}(\nu)\right)' T_n \left(D_n^{-1}\mathbf{h}(\nu)\right), \tag{29}$$

where $\mathbf{h}(\nu) = T_n^{-1}\mathbf{f}(\nu)$. It is important to note that the elements of $\mathbf{h}(\nu)$, say $h_r(\nu)$, $r = 1, \dots, k_n$, are linear splines in ν whose coefficients are given by the rth row of T_n^{-1}. The elements of T_n are defined in (28); the results of Graybill [(1983), page 286] for tridiagonal symmetric Toeplitz matrices may be used to find the elements of T_n^{-1}. Asymptotically, the elements form a geometric sequence starting on the diagonal with the value $a = 2\sqrt{3}/3$. For any element in the same row falling p columns away from the diagonal element (to the left or to the right), the value of the sequence is $a\eta^p$, where $\eta = \sqrt{3} - 2$. These results are typical of the geometric falloff away from the diagonal discussed in Lemma A.1.

Since the elements of T_n^{-1} are decreasing exponentially fast to 0 and $\mathbf{f}(\nu)$ is nonzero only for $f_{g_\nu - 1}(\nu)$ and $f_{g_\nu}(\nu)$ only, the elements of $\mathbf{h}(\nu)$ decrease exponentially fast to 0 away from the index g_ν. Returning to (29), this implies that the influence of D_n^{-1} on $(D_n^{-1}\mathbf{h}(\nu))'T_n(D_n^{-1}\mathbf{h}(\nu))$ is minimal except in a neighborhood about ν. Thus, by replacing D_n^{-1} with $\operatorname{diag}(1/\alpha(\nu) \cdots 1/\alpha(\nu))$, it is easy to see that

$$\begin{aligned}\sigma^2(\nu, \pi_n) &\approx \alpha(\nu)\big(D_n^{-1}\mathbf{h}(\nu)\big)'T_n\big(D_n^{-1}\mathbf{h}(\nu)\big)\\ &\approx \frac{1}{\alpha(\nu)}\mathbf{h}'(\nu)T_n\mathbf{h}(\nu)\\ &= \frac{1}{\alpha(\nu)}\mathbf{f}'(\nu)T_n^{-1}\mathbf{f}(\nu),\end{aligned}$$

where the last step follows from the definition of $\mathbf{h}(\nu)$.

To complete the derivation of $\sigma^2(\nu, \pi_n)$, we need to determine the value of $\mathbf{f}'(\nu)T_n^{-1}\mathbf{f}(\nu)$. Let us define $\pi_n = (\nu - \tau_{g_\nu})/\Delta_n$, the relative position of ν between its two bracketing knots for a given n. Then, using the definitions of the linear B-spline basis functions $f_j(\nu)$, $j = 1, \dots, k_n$, some elementary algebra yields that

$$\begin{aligned}\mathbf{f}'(\nu)T_n^{-1}\mathbf{f}(\nu) &= a(1 - \pi_n)^2 + 2\pi_n(1 - \pi_n)a\eta + a\pi_n^2\\ &= (4\sqrt{3} - 4)\pi_n^2 - (4\sqrt{3} - 4)\pi_n + \frac{2\sqrt{3}}{3}.\end{aligned}$$

Thus, for values of ν away from the boundaries of Θ and $n \gg 1$, we have that

$$\sigma^2(\nu, \pi_n) \approx \frac{1}{\alpha(\nu)}\left[(4\sqrt{3} - 4)\pi_n^2 - (4\sqrt{3} - 4)\pi_n + \frac{2\sqrt{3}}{3}\right].$$

For $\nu \in \Theta^o$, it is obvious from this expression that $\sigma^2(\nu, \pi_n)$ is a quadratic function in π_n. In addition, taken as a function of π_n, $\sigma^2(\nu, \pi_n)$ is minimized at $\pi_n = 0.5$. The fluctuation in $\sigma^2(\nu, \pi_n)$ for a given ν is bounded; in fact, for large n, the ratio of the maximum to the minimum values of $\sigma^2(\nu, \pi_n)$ for a given ν over a given interval is exactly $\sqrt{3} + 1$.

Initially, this result may appear curious; however, since one can think of fitting the hazard model (1) as doing something akin to many local linear regressions, it is a sensible result. The pointwise confidence bands in linear regression are tightest near the mean value, and flare out from each point; a similar phenomenon appears to be happening here. □

Acknowledgments. The authors wish to thank the referees for a very careful reading of this manuscript and their helpful comments, both of which contributed to a greatly improved presentation of our results.

REFERENCES

AALEN, O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6** 701–726.

ANDERSEN, P. K. and GILL, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10** 1100–1120.

APOSTOL, T. M. (1957). *Mathematical Analysis*. Addison-Wesley, Reading, MA.

BRILLINGER, D. R. (1980). *Time Series: Data Analysis and Theory*. Holden-Day, San Francisco.

CHUNG, K. L. (1974). *A Course in Probability Theory*, 2nd ed. Academic Press, New York.

DE BOOR, C. (1978). *A Practical Guide to Splines*. Springer, New York.

DEMKO, S. (1977). Inverses of band matrices and local convergence of spline projections. *SIAM J. Numer. Anal.* **14** 616–619.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications*, 2nd ed. Wiley, New York.

FLEMING, T. R. and HARRINGTON, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley, New York.

FOUTZ, R. V. (1977). On the unique consistent solution to the likelihood equations. *J. Amer. Statist. Assoc.* **72** 147–148.

GOLUB, G. and VAN LOAN, C. F. (1989). *Matrix Computations*. Johns Hopkins Univ. Press.

GRAYBILL, F. A. (1983). *Matrices with Applications in Statistics*, 2nd ed. Wadsworth, Belmont, CA.

HJORT, N. L. (1992). On inference in parametric survival data models. *Internat. Statist. Rev.* **60** 355–387.

KALBFLEISCH, J. D. and PRENTICE, R. L. (1980). *The Statistical Analysis of Failure Time Data*. Wiley, New York.

KOOPERBERG, C. K., STONE, C. J. and TRUONG, Y. K. (1995a). Hazard regression. *J. Amer. Statist. Assoc.* **90** 78–94.

KOOPERBERG, C. K., STONE, C. J. and TRUONG, Y. K. (1995b). The L_2 rate of convergence for hazard regression. *Scand. J. Statist.* **22** 143–158.

PORTNOY, S. (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Ann. Statist.* **16** 356–366.

RUDIN, W. (1964). *Principles of Mathematical Analysis*. McGraw-Hill, New York.

SCHUMAKER, L. (1981). *Spline Functions: Basic Theory*. Wiley, New York.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

SPECKMAN, P. (1985). Spline smoothing and optimal rates of convergence in nonparametric regression models. *Ann. Statist.* **13** 970–983.

STONE, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360.

STONE, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053.

STONE, C. J. (1986). The dimensionality reduction principle for generalized additive models. *Ann. Statist.* **14** 590–606.

STONE, C. J. (1990). Large-sample inference for log-spline models. *Ann. Statist.* **18** 717–741.
STONE, C. J. (1991). Asymptotics for doubly flexible log-spline response models. *Ann. Statist.* **19** 1832–1854.
STONE, C. J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation (with discussion). *Ann. Statist.* **22** 118–184.
STRAWDERMAN, R. L. (1992). Statistical methods in the surrogate marker problem. Sc.D. dissertation, Harvard School of Public Health.
STRAWDERMAN, R. L. and TSIATIS, A. A. (1995). On consistency in parameter spaces of expanding dimension: an application of the inverse function theorem. *Statist. Sinica*. To appear.

DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF MICHIGAN
1420 WASHINGTON HEIGHTS
ANN ARBOR, MICHIGAN 48109-2029

DEPARTMENT OF BIOSTATISTICS
HARVARD SCHOOL OF PUBLIC HEALTH
677 HUNTINGTON AVENUE
BOSTON, MASSACHUSETTS 02115